SEARCHING OPEN DATA SETS

Table of Contents

U.S. Census Bureau Data Sources	
Demographic Surveys	
Economic Surveys	
American Fact Finder	
Accessing AFF	
Community Facts Search	
EDGAR	12
U.S. Bureau of Economic Analysis (BEA)	13
Organization for Economic Cooperation and Development (OECD)	17
World Bank	21
UNDATA	25
Data.gov	28
Browse by Subject	29
Browse by Location	32
References	3:

U.S. Census Bureau Data Sources

Data collection at the Census Bureau is divided into two primary categories: (1) Demographic surveys and (2) Economic surveys. The bureau also collects information in other specialized surveys that are developed with strategic partners (United States Census Bureau, 2015).

These surveys allow the Census Bureau to collect and analyze social, economic, and geographic data.

Demographic Surveys

The 5 primary household surveys conducted by the Census Bureau are (United States Census Bureau, 2015):

2015).			
Survey	Focus	Schedule	
American Community Survey (ACS)	The ACS is a mandatory, ongoing statistical survey that samples a small percentage of the population giving communities the information they need to plan investments and services.	Yearly	
Survey of Income and Program Participation (SIPP)	SIPP collects data and measures change for many topics including: economic wellbeing, family dynamics, education, assets, health insurance, childcare, and food security. Linked to SSA disability and earnings data.	Designed as a continuous series of national panels. Each panel features a nationally representative sample interviewed over a multi-year period lasting approximately four years.	
Current Population Survey (CPS)	Sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), is the primary source of labor force statistics for the population of the United States.	Monthly	

Consumer Expenditure Survey (CE) The survey data are collected for the Bureau of Labor Statistics by the U.S. Census Bureau. Data comes from two surveys, the Quarterly Interview Survey and the Diary Survey, that provide information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and

Quarterly Interview survey and the Diary survey.

Economic Surveys

The Census Bureau collects economic data on nine economic sectors (United States Census Bureau, 2015):

- Multi-Sector (Programs covering several sectors)
- Construction (Buildings, alterations and public works)
- Governments (Local, state and federal agencies)
- Foreign Trade (Exports, imports and participants)
- Manufacturing (Companies, operations and shipments, products)
- Mining (Minerals, gases and initial processing)
- Retail (Merchandise for personal or home use)
- Services (Personal, business and transport services)
- Wholesale (Merchandise for business use)

American Fact Finder

There are many methods for viewing and downing Census Bureau data, I will be focusing on American Fact Finder (AFF).

I am only covering the basics. I can conducts tutorials/webinars on specific topics, upon request.

American Factfinder (AFF) is a search tool that provides access to data about the United States, Puerto Rico, and the Island Areas.

The following data are available on American FactFinder:

- American Community Survey
- American Housing Survey

- Annual Economic Surveys
- Annual Surveys of Governments
- Census of Governments
- Decennial Census
- Economic Census
- Equal Employment Opportunity (EEO) Tabulation
- Population Estimates Program
- Puerto Rico Community Survey

There are three primary options for accessing AFF data:

Community Facts: Find popular facts (population, income, etc.) and frequently requested data about a geographic area.

Guided Search: Finds Census data by specifying search criteria using a systematic approach. The following options are available via Guided Search:

- Searching for information about people (e.g. age, sex, income, poverty, or education).
- Searching for information about housing (e.g. housing units, household type, and value of home). Searching for information about businesses or industries (e.g. annual payroll, sales, receipts, or number of employees).
- Searching for information from a specific dataset (e.g. 2006-2010 American Community Survey 5-Year Estimates, or the 2010 Census Summary File 1).
- Searching for a specific table number or a table title (e.g. P1, EC0700A1, b04*, Social Characteristics).

Advanced Search: Table selection by topic clusters.

Downloading Data: Available methods for downloading data include:

- Download from the search results page: Download up to 10 tables from the search results page in a comma-delimited format (.csv). This format can be imported into Microsoft Excel, as well as other spreadsheet and database programs.
- Download when viewing a table: Download a single table from the table view in commadelimited (.csv), PDF, Microsoft Excel (.xls), or Rich Text Format (.rtf) formats.
- Download spatial data from reference maps and thematic maps.
- Download Shapefiles: can be imported into various geographic information systems (GIS) software packages.

Accessing AFF

 From the Census Bureau website, choose Data>Data Tools and Apps>American FactFinder (Figure 1).



Figure 1: Accessing AFF data.

2. From the AFF homepage, click on Advanced Search (Figure 2).



Figure 2: AFF homepage.

Community Facts Search

- 1. Click on Community Facts to open up the search box.
- 2. Type Bethesda in the Quick Search Box.
- 3. As you can see in Figure 3, AFF will auto-complete based on Census geographic place names. CDP stands for Census Designated Place.



Figure 3: Searching Community Facts Search.

- 4. Select **Bethesda**, and click **Go**.
- 5. When you click Go, you get a quick snapshot panel that provides basic information about Bethesda
- 6. There ae three options from dropdown menu: (1) 2010 Census Total Population, (2) 2014 Population Estimates, or (3) 2013 ACS 5-year Population Estimates.
- 7. AFF also lists commonly requested tables in the Quick Snapshot panel.
- 8. From the 2013 American Community Survey section, click on *Demographic and Housing Estimates* (Figure 4).

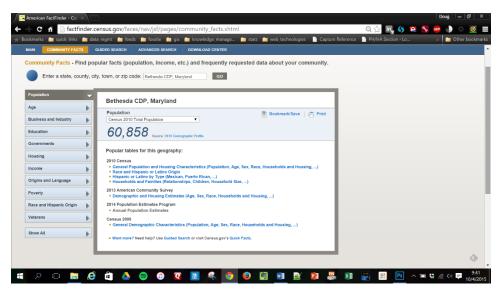


Figure 4: Bethesda Community Facts results panel.

9. A table will open up that displays more descriptive information about Bethesda (pop for males and females), and a breakdown of population by various cohorts (Figure 5).

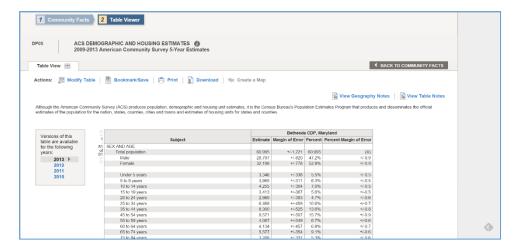


Figure 5: ACS Demographics and Housing Estimates results.

- 10. Click *Main* (top navigation), to go back to go back to the main page.
- 11. Scroll down the page to the What We Provide section (great for users that are new to AFF).
- 12. Scroll past of the **following data are available on American FactFinder** section.
- 13. Click on the *More* Link (lower right side of the page).
- 14. As you can see in Figure 6, this page is organized by Census/Surveys, Downloading Data, Thematic Maps, and Reference Maps.

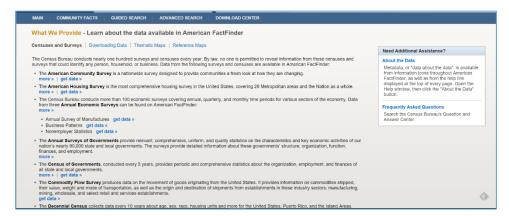


Figure 6: AFF What We Do page.

- 15. Click *Main* (top navigation), to go back to go back to the main page.
- 16. Scroll down to the **EEO tabulations** section.
- 17. The **More** links generally take you to a page that provides an overview of the data set, how the data is collected, and any limitations related to its use or analysis.
- 18. If you clicked on *Get Data*, AFF would open to the search page...with the EEO Tabulation data preselected.
- 19. Click on the *EEO Tabulation Data* link, to get the actual data (Figure 7).



Figure 7: EEOC data page.

- 20. As you can see in Figure 8, there are 169 tables.
- 21. From the **Refine Your Search** box, you can refine your search by: (1) Table or Topic name (if you click on the radio boxes under the search box, you can limit the search to race/ancestry, or industries, or occupations), or by (2) state, county or place.
- 22. If available, you can filter the list of tables, using the **Show results from** drop-down menus (Figure 8).

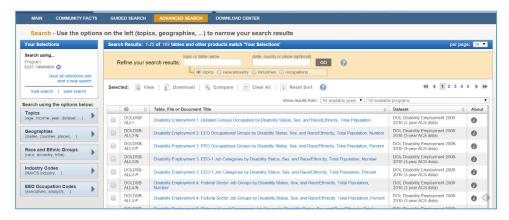


Figure 8: EEOC tables page.

- 23. A key feature of the results page is the **About** icon.
 - a. Click on the **About** icon for the following table: DOLDSB-ALL1.
 - b. As shown in Figure 9, this page provides information about the program, the survey, and the dataset that was used to generate the table. Also vital to understanding the data are the five dimensions of the table: (1) Geography, (2) Occupation Code, (3) Boxhead, (4) Estimate, and (5) Subject.

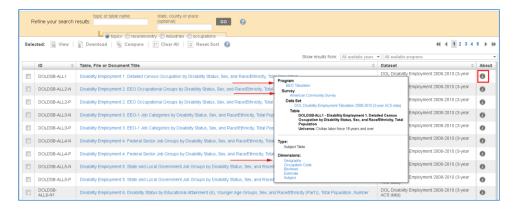


Figure 9: About information from the DOLDSB-ALL1 Table.

- c. Click on the Occupation Code to see what happens.
- d. As you can see in Figure 10, all of the EEOC Occupation Codes are listed.
- e. Scroll down to, and click on, the **Social and community service managers 0420 (SOC 11-9151)** code (Figure 10).

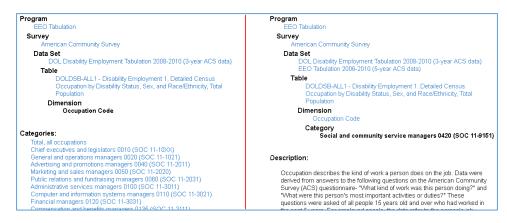


Figure 10: AFF Dimension – Occupation Code – 11-9151.

- f. Close this window to return to the list of tables.
- 24. Click on the DOLDSB-ALL1 Table.
- 25. This opens up the Table Viewer (Figure 11).
- 26. Most tables in AFF can be customized using the Modify Table feature (Figure 11).
- 27. The Actions section of this view displays the options for interacting with table data (Figure 11).
- 28. The View Geographic Notes, and View Table Notes section of this view provides information about the table data (Figure 11).
- 29. From the Estimate dropdown, you can choose Estimate or Margin of Error (Figure 11).
- 30. You cannot change anything from the Geography drop because the table was created with no selected geography. If the table is available for the United States, then the default geography will be the United States.
- 31. If you want to add a particular geography, click on Add/Remove Geographies (Figure 11).

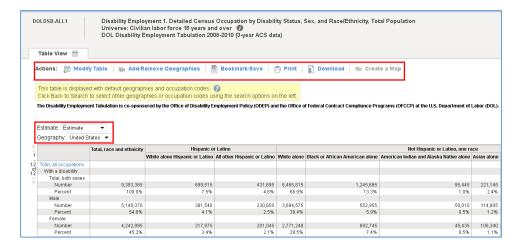


Figure 11: AFF Table Viewer.

- 32. Click on Add/Remove Geographies.
- 33. Search for Washington, DC. (As you start typing AFF will fill in the search box, based on geolocations).
- 34. Select Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area as the geography.
- 35. Click Go (Figure 12).
- 36. Click Show Table (Figure 12).

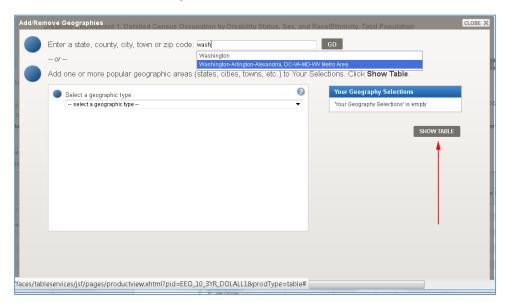


Figure 12: Changing Geographies in AFF.

1. Figure 13, shows the table with the updated geographies.

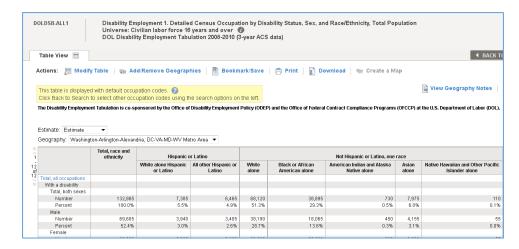


Figure 13: AFF table with updated geographies.

If elements of this table were "mappable" then the Create a Map link would be active. I talk about mapping AFF data in my Mapping Health Data webinar.

EDGAR

Every public company in the United States, meaning companies that offer stocks on exchanges must file certain documents with the SEC (United States Securites and Exchange Commisson, 2015).

The SEC's <u>EDGAR database</u> provides free public access to corporate information, allowing you to quickly research a company's financial information and operations by reviewing registration statements, prospectuses and periodic reports filed on Forms 10-K and 10-Q (Figure 14) (United States Securites and Exchange Commisson, 2015).



Figure 14: SEC EDGAR homepage.

If you are new to this system, I recommend that you read the <u>Researching Public Companies through EDGAR: A Guide for Investors.</u>

1. Click on Company Filings (left-hand section of the page).

2. This brings up the EDGAR search form. If you click on More Options, the advanced search fields are displayed (Figure 15).

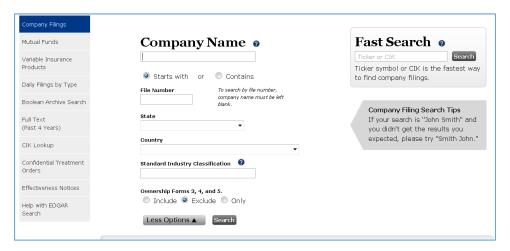


Figure 15: Company filings search page.

- 3. Type in the name **Google** and click **search**.
- 4. Figure 16 shows the results of the search. You will also notice that the results also list the Standard Industrial Classification (SIC) code for Google (7370).

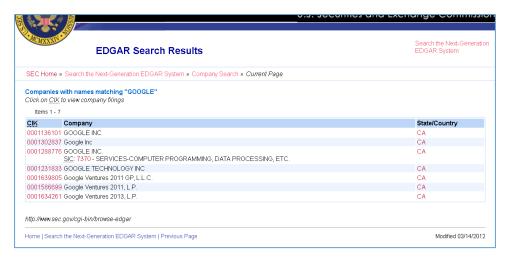


Figure 16: Company filings search results.

- Click on the following entry: GOOGLE INC. SIC: 7370 SERVICES-COMPUTER PROGRAMMING, DATA PROCESSING, ETC.
- 6. As you scroll down the results page, you will notice that some documents (Q-10 form) are interactive (Figure 17).
- 7. If you do not see the Q-10 forms, you can enter Q-10 in the Filling Type filter box (Figure 17).



Figure 17: Google filings results page.

Interactive forms allow you filter any report or export the data to Excel (Figure 18).

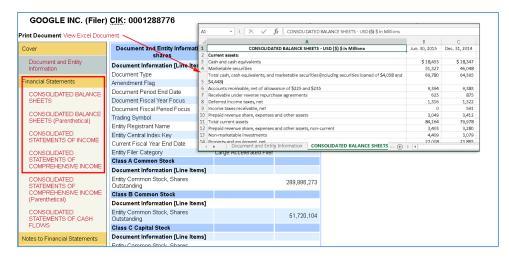


Figure 18: Google table in interactive mode.

U.S. Bureau of Economic Analysis (BEA)

The BEA is an agency of the Department of Commerce. Along with the Census Bureau, BEA is part of the Department's Economics and Statistics Administration.

BEA produces economic accounts <u>statistics</u> that enable government and business decision-makers, researchers, and the American public to follow and understand the performance of the Nation's economy (US Department of Commerce, 2015).

- 1. Click on Interactive Data tab.
- 2. Click on Industry Data>GDP by Industry (Figure 19).

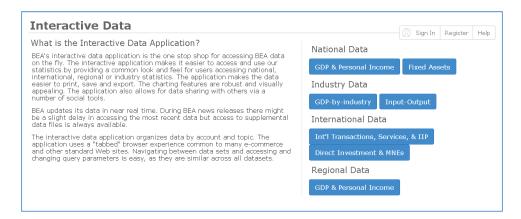


Figure 19: Interactive Data page.

- 3. Figure 20 displays the **GDP-by-industry** search page. You can always tell which set of data you are searching, because the name is highlighted in orange.
- 4. Click on Begin Using the Data (Figure 20).

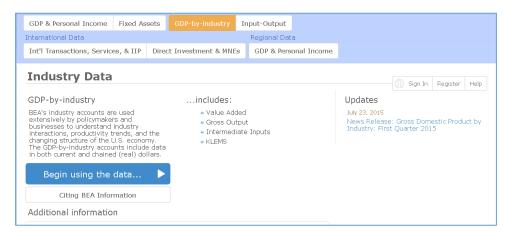


Figure 20. Industry Data page.

5. We want to search for KLMES data (K-capital, L-labor, E-energy, M-materials, and S-purchased services). So click on the arrow next the KLEM category (Figure 21).

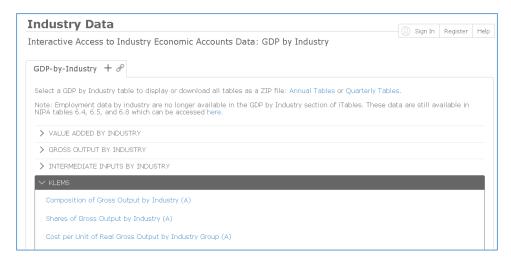


Figure 21: BEA Interactive Chart.

- 6. We are interested in the Shares of Gross Output by Industry (A), so click on that value.
- 7. As illustrated by Figure 22, you have lots of options for refining the data table or outputting the data as a chart.
 - a. Modify: Allows you customize the table by year or by industry (code list).
 - b. Download: in Excel, CSV, or PDF format.
 - c. Chart: Create interactive charts.

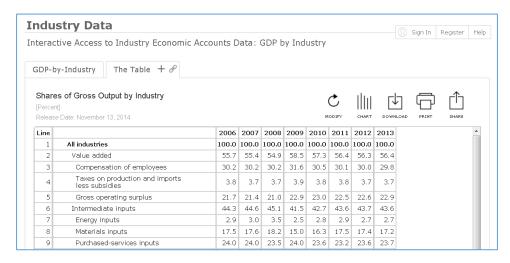


Figure 22: Industry Data results page.

- 8. Click Chart.
- 9. This brings up the chart selection screen (Figure 23).

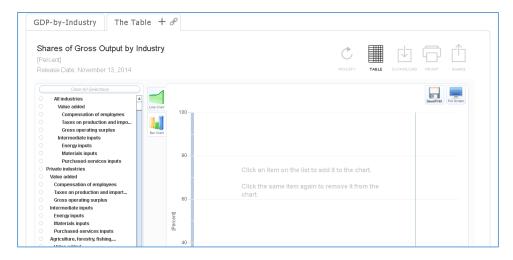


Figure 23 BEA data chart selection screen:

- 10. Click the Full Screen option.
- 11. Select each input under All Industries>Intermediate inputs (Figure 24).
- 12. This will add the energy input, materials inputs, and purchased-services inputs on the chart (Figure 24).

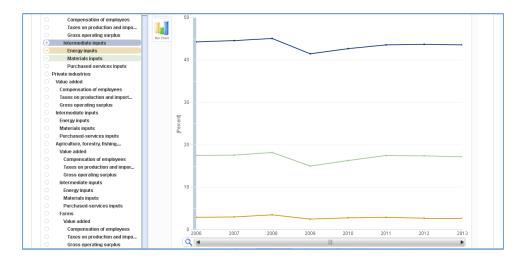


Figure 24:

- 13. You have two options for the BEA charts: (1) line chart, or (2) bar chart.
- 14. Scroll up and click on Full Screen mode (Figure 25).
- 15. It is now easier to see the data marker (Figure 25).
- 16. The data marker allows you to hover over a data point, to see its actual value.

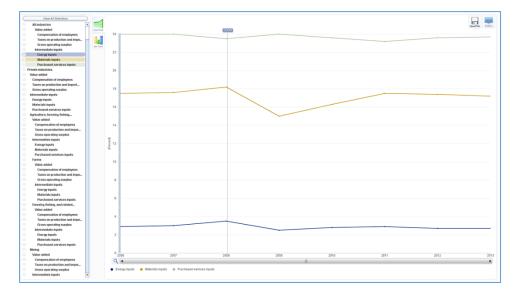


Figure 25: BEA chart in full-screen mode.

- 17. The data marker also allows you to zoom in on a range of data points.
- 18. Place the data marker over the year 2008.
- 19. Press down on your mouse a drag the data marker from 2008 to 2010.

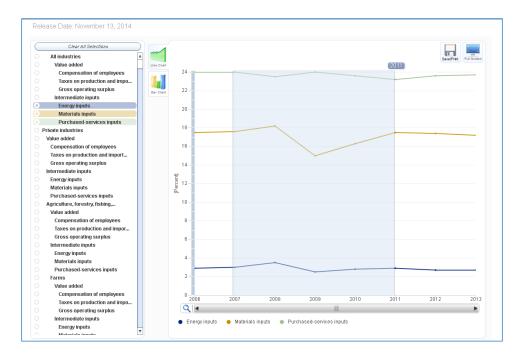


Figure 26: Selecting a range of data from a BEA chart.

- 20. Release your mouse, and the table will refresh based on your selection.
- 21. If you select Save/Print, the chart will be exported as a table.

Organization for Economic Cooperation and Development (OECD)

The OECD is an international organization that promotes policies to improve the economic and social web-being of people (Organization for Economic, 2015).

- 1. From the OECD homepage, click on the Data tab.
- 2. As illustrated in Figure 27, you can search the Catalogue of OECD databases, or browse by topic or browse by country.
- 3. Scroll down to the Statistical Resources section of the data page (Figure 27).
- 4. Here you can find options for creating custom queries, browse a list of statistical tools, or find out more information about the OECD API. The statistical tools section is primarily a list of resources. However, you can access the OECD.Stat from this page. OECD.Stat includes data and metadata for OECD countries and selected non-member economies.



Figure 27: OECD data page.

5. Click on the Topic pulldown to bring up the topics page (Figure 28).



Figure 28: OECD topics page.

- 6. This will bring you to the OECD Energy page. Scroll down to the Renewable Energy section. As you scroll down the page you will notice that some data elements are reports, some are indicators, and some are links.
- 7. Click on the Renewable Energy link. This will bring you to the Energy Indicator data page (Figure 29).
- 8. The chart that is displayed provides information about the Total, % of total energy generation, 2002 2012.

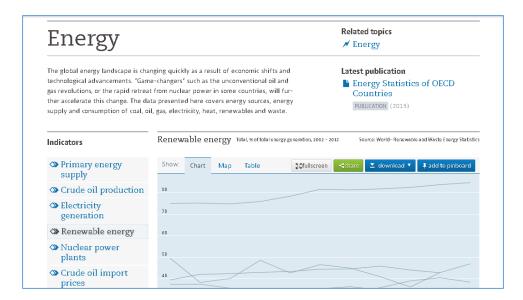


Figure 29: OECD Renewable Energy chart.

- 9. If you scroll down past the table you will see options for customizing the chart.
- 10. Select Highlight Countries.
- 11. Check the boxes for Austria, Australia, Belgium, Brazil, and Iceland (Figure 30).
- 12. Click None, under the Select Background section to remove all of the other data points.
- 13. Click Close.



Figure 30: Filtering an OECD Renewable Energy chart.

14. Scroll back to the top of the page, and select Full Screen (Figure 31).

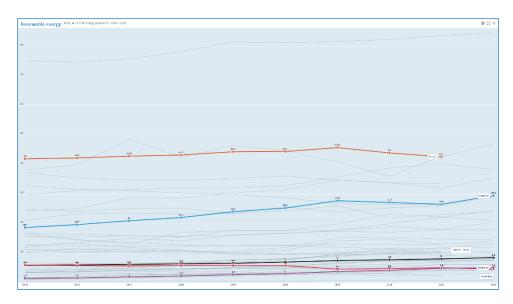


Figure 31: OECD Renewable Energy chart in full screen mode.

- 15. If you mouse over one of the trend lines, you will see the data points for the entire line (Figure 32).
- 16. This makes it easier to compare the data from Iceland with the OECD baseline data (black line in the chart).



Figure 32: Comparing values for OECD data.

17. Options at the top of the chart include; (1) Sharing the data, (2) downloading the data, or (3) pinning the data in an external website.

World Bank

The World Bank provides access to international development data. Data at the World Bank is organized by: Country, Topic, Economic Indicators, Data Catalog, and Microdata (Figure 33).

The Data Catalog provides access to all available World Bank datasets, including databases, preformatted tables, reports, and other resources (World Bank, 2015).

DataBank is an analysis and visualization tool that contains collections of time series data on a variety of topics. (World Bank, 2015) Using DataBank you can create your own queries; generate tables, charts, and maps; and easily save, embed, and share them (World Bank, 2015).

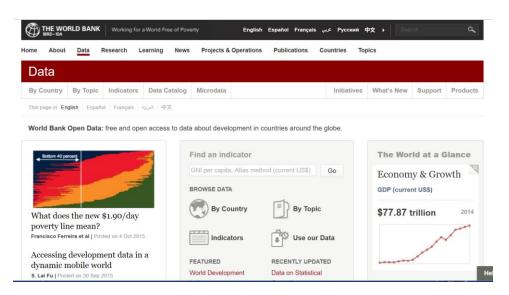


Figure 33: World Bank data page.

The World at a Glance section provides information on the economy and growth, health, education, and Environment.

- 1. Type Poverty in the **Find an Indicator** search box (Figure 34).
- 2. Select Poverty gap at national poverty lines (%), and click Go.

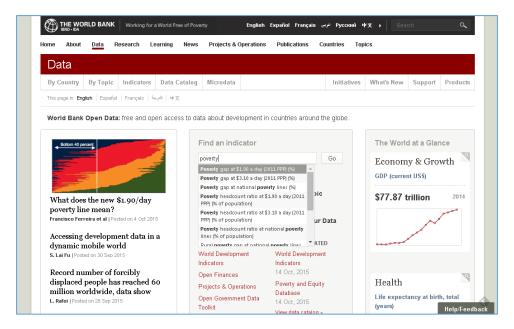


Figure 34: Searching World Bank data using World Bank indicators.

- 1. Figure 35, display the results for the **Poverty gap at national poverty lines** data. There are a lot of options displayed in Figure 35:
 - a. At the top of the page is a description of how this percentage is calculated, the license for the data, and source.
 - b. If you click View in WDI Tables, you can see the table used to generate this chart (Figure 36).
 - c. You can also view this data as a (1) map, or (2) graph.
 - d. If you click on the scissors near the top of the table, you can embed this table in a webpage.
 - e. You can look at other indicators by selecting them from the Featured Indictors dropdown menu.

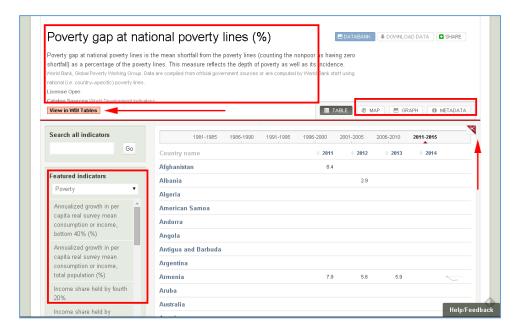


Figure 35: Poverty Gap results screen.

2. Right-click on the View in WDI Tables link to open the table in a new tab (Figure 36).

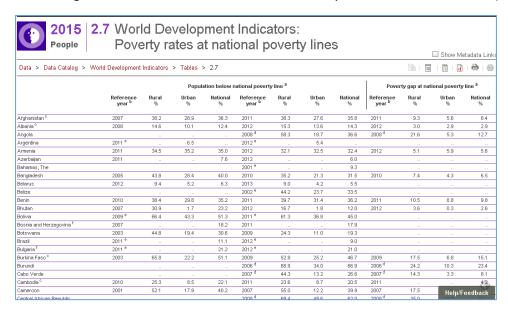


Figure 36: The WDI (World Development Indicators) table for this data.

- 3. The Download Data link is near the top of the page. This allows me to download the data as Excel, CSV, or XML format.
- 4. If you click on the *DataBank* link, this opens up the data in *DataBank* catalog (Figure 37).

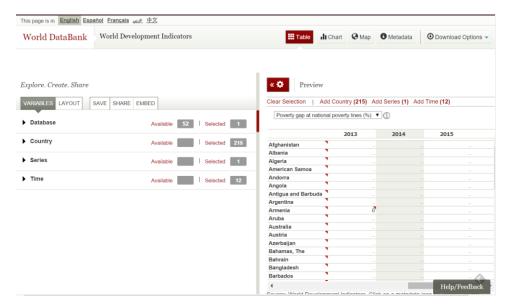


Figure 37: World Bank data in view in the DataBank.

- 5. This view provides you with a number of options for refining, for example adding new variables to your query or changing the layout of the table.
- 6. Click on the Home icon to go back the World Bank homepage.
- 7. Click on Data>Data Catalog.
- 8. The Data Catalog provides access to all available World Bank datasets, including databases, preformatted tables, reports, and other resources (Figure 37).
- 9. You can search the Data Catalog, download the entire catalog, or connect to the data via the Data Catalog API.

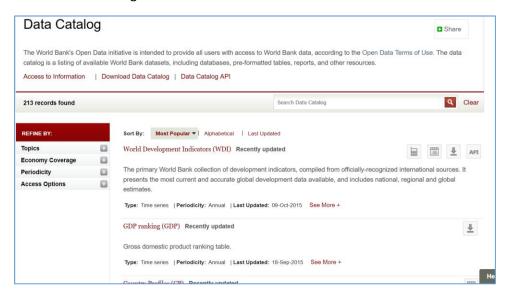


Figure 37: World Bank Data Catalog.

10. If you download the Data Catalog, you are not downloading the entire dataset. You are downloading a list of the resources that are available, their URL, data type, and imitations on use.

11. Figure 38 is the Data Catalog, which I have opened in Microsoft Excel.

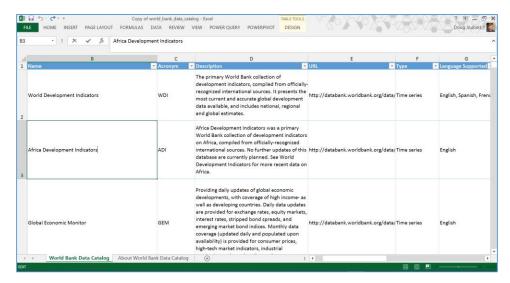


Figure 38: Data Catalog, viewed in Excel.

UNDATA

The United Nations provides a forum for its 193 member states. As part of its mission, the UN provides access to its own data (United Nations Statistics Division, 2015).

If you go to the main UN website, and scroll down to the bottom of the page...you can see a list of their

UNDATA provides access to the UN statistical databases through a single <u>entry point</u>, where users can search and download a variety of statistical resources of the UN system (United Nations Statistics Division, 2015).

There are a number of features on the homepage (Figure 40).

- You can perform a keyword search.
- You can browse by category, using one of the six UN glossaries.
- Population Estimates and Projections, (4) SDMX [Statistical Data and Metadata eXchange], (5) National Accounts Main Aggregates, or (6) System of National Accounts 1993.
- Browse by Database, or by Country Data Services (this option also includes a country profile page).
- Build reports from the MBS (Monthly Bulletin of Statistics) resources.



Figure 40: UNDATA homepage.

Figure 40: UNDATA homepage.

- 1. Click on Crime>UNODC Homicide Statistics 2012, UNODC, from the Databases section.
- 2. This will bring you to the appropriate section on the UNDATA Databases page (Figure 41).
- 3. Click on Preview, if you want to preview the data.
- 4. Click on View Data (Figure 41).



Figure 41: Homicide statistics entry.

5. Figure 42 displays the options for this dataset.

- 6. The first thing I always do when I am working with a new dataset...is scroll to the bottom of the page. This will allow you to see the data **source types**. In this case the CJ: Criminal Justice and PH: Public Health.
- 7. You will also be able to see the 14 datasets that contribute to this table. When you look at the various data sources you will see that this data comes from non-UN bodies. For example, PAHO and the World Health Organization.

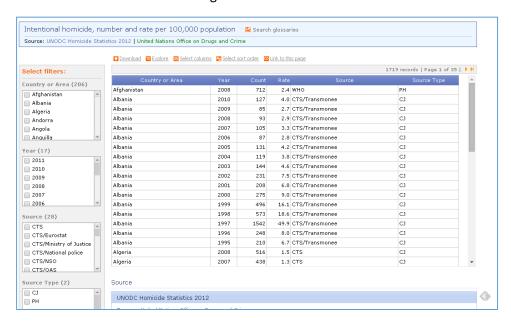


Figure 42: UNODC Homicide Statistics 2012 data page.

8. Select the Select Columns option near the top of the table. This option allows you to select particular columns to display on the table (Figure 43).

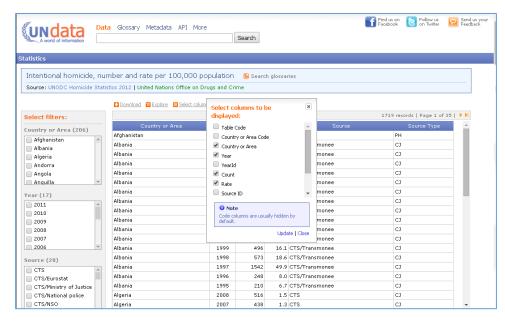


Figure 43: Selecting columns on a UNDATA table.

- 9. Click the UNDATA icon to go back to the main page.
- 10. From Country Data Services, go down to Austria, and select the Country Profile icon.
- 11. Figure 44 is displaying the country profile for Austria.

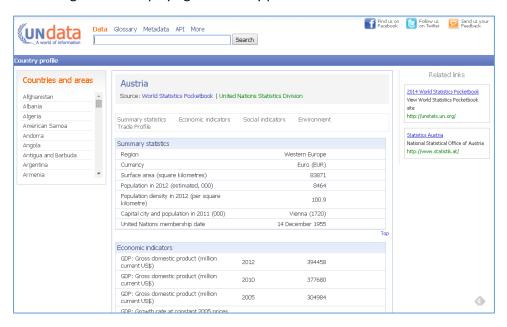


Figure 44: Austria country profile.

- 12. Click the UNDATA icon to go back to the UNDATA homepage.
- 13. Please note that if I had clicked on the Austria name (rather than the country profile icon), I would have gone to an external (non-UN) data site. I mention this because these sites are managed externally...so you might get a broken link.

Data.gov

The <u>2013 Federal Open Data Policy</u> requires that newly-generated government data be made available in open, machine-readable formats (U.S. General Services Administration, 2015).

Data.gov is primarily a federal open government data site. However, state, local, and tribal governments can also syndicate their metadata (U.S. General Services Administration, 2015).

Data.gov does not host data directly, but rather aggregates metadata about open data resources in one centralized location (U.S. General Services Administration, 2015).

You can search Data.gov from its catalog of government data from across the Federal Government. Once in the catalog, to find datasets you can:

- 1. Enter keywords in the search box.
- 2. Browse on the left side through types, tags, formats, groups, organization types, organizations, and categories. Clicking on multiple items narrows your search. You can click on the "x" to the side of any single item to remove it from the search, or "clear all" to remove all selected items in a category.

- 3. Search by geospatial area by drawing a boundary box on the map at the left side and clicking "Apply" to find all datasets that are tagged for that geographic area.
- 4. Once you find a dataset or tool of interest, click on the title and you will be taken to a page with more details on that specific dataset or tool. Some datasets are downloadable, while others are links to web sites or apps that help you access or use the data.

Please note that by accessing datasets or tools offered on Data.gov, you agree to the Data Policy, which you should read before accessing any data. If there are additional datasets that you would like to see included on this site, please suggest more datasets here.

Browse by Subject

1. Select Science and Research from the Topics pulldown menu (Figure 45).

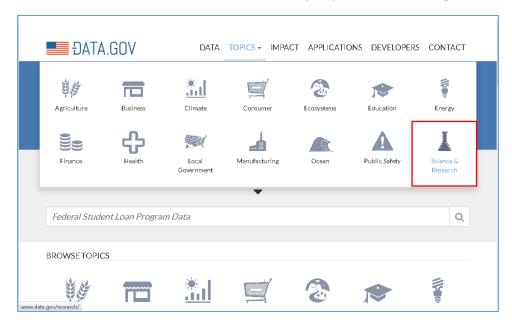


Figure 45: Browsing by Topic on Data.gov

2. From the Science and Research page, click on the Data Tab (Figure 46).

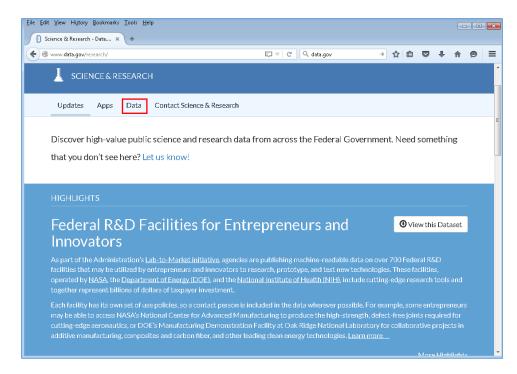


Figure 46: Data tab from Data.gov

- 3. As illustrated by Figure xx, there are 169 datasets tagged as "research" (Figure 47).
- 4. Data can be sorted by: relevance, name, last modified, popular, or date uploaded.
- 5. Starting from the top of Figure 47, you can:
 - a. Filter by location: type in a geo-location or, click on the edit icon to draw a region on the map. Please note that this works better when searching...rather than when filtering.
 - b. Add a topic area by clicking on a topic.
 - c. Note that all of the data on this page is from a federal agency.
 - d. Note the file format for the actual data.

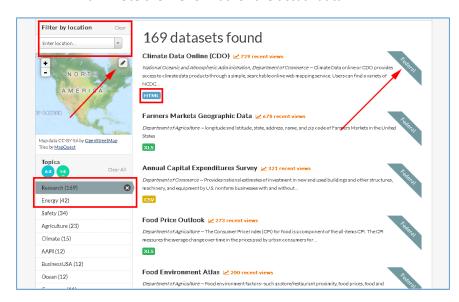


Figure 47: Research datasets in Data.gov

Browse by Location

- 1. Go back to the Data.gov homepage and click Clear to remove any previously defined filters.
- 2. Type in Seattle, WA in the Filter by Location search box (Figure 48).
- 3. Unlike AFF, you cannot select multiple zip codes (which is how the data is geo-coded).
- 4. Pick the first zip code in the list (98101).

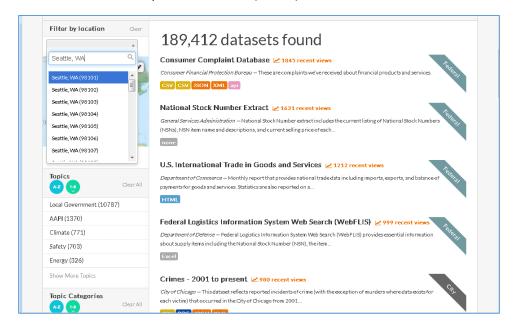


Figure 48: Searching by location in Data.gov.

- 5. As you can see in Figure 49, the data has been filtered by location.
- 6. Click on the map edit icon.

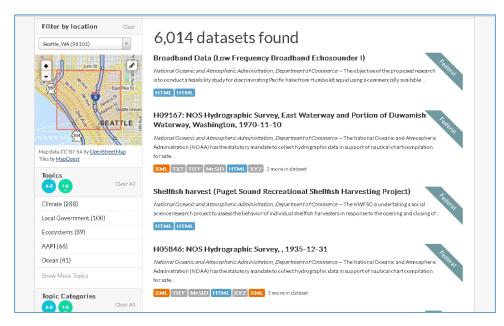


Figure 49: Data.gov data filtered by location.

- 7. This will make the map bigger so it is easier to select a different location (Figure 50).
- 8. Use the symbol to zoom out on the map.

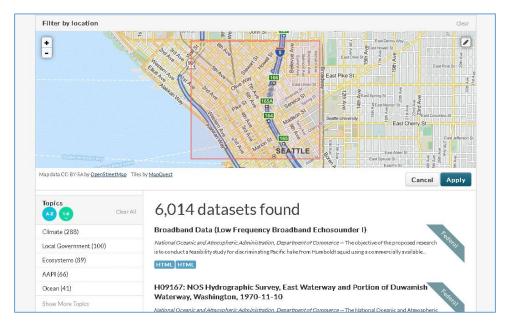


Figure 50: Refining a location search in Data.gov

- 9. Zoom out far enough to select the city of Bellevue.
- 10. Use the select tool to select a region in Bellevue.
- 11. Click Apply to refresh the data (Figure 51).

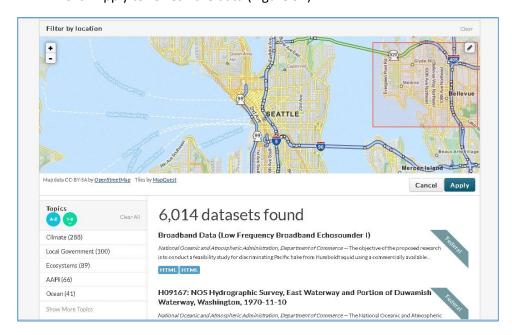


Figure 51: Refining a location search in Data.gov

The disadvantage of using location is that there is no pan function on the map.

References

Organization for Economic, D. (2015). OECD Data. Retrieved from https://data.oecd.org/ U.S. General Services Administration. (2015). About Data.gov - Data.gov. Retrieved from https://www.data.gov/about

United Nations Statistics Division. (2015). UNdata. Retrieved from http://data.un.org/Host.aspx?Content=About

United States Census Bureau. (2015). Data collection at the U.S. Census Bureau.

United States Securites and Exchange Commisson. (2015). EDGAR Search Tools. Retrieved from https://www.sec.gov/edgar/searchedgar/webusers.htm

US Department of Commerce, B., Bureau of Economic Analysis. (2015). Bureau of Economic Analysis. Retrieved from http://www.bea.gov/about/mission.htm

World Bank. (2015). The World Bank DataBank. Retrieved from http://databank.worldbank.org/data/home.aspx